IFAP and Migration Databases

Processing Summary

Preliminary Steps

The IFAP Notes databases need to be prepared for the actual bulk load processing. For example, every IFAP database needs an additional view, All Docs, which contains every document in the database. Agents in the Metrics & Migration database will use this view extensively.

Several other preliminary steps are needed to prepare the Migration databases. For example, special views have to be built in Regulations Compilations and Federal Registers to extract linked information.

Bulk Processing

For convenience, the 39 IFAP Notes' databases are divided into 8 migration databases. Each migration database has all the agents needed to process the data in the database. During bulk processing, data is copied from the original IFAP databases into the migration databases. We refer to the migration databases as AFIS 1 through AFIS 8. ASIF is short 'All Fields In Sequence'. The AFIS databases will contain all data, sequenced correctly.

In the migration databases, the data are massaged, checked for errors, combined in various ways, and written to Rich Text files. The Rich Text files are later converted to HTML files. Attachments such as PDF and DOC files are also written to the standard directory and detached from the original Notes documents.

As a final step in each Migration (AFIS) database, Metadata files for are built as the final step in basic processing. These contain data for every document processed. An example of metadata is Publication Date or Award Year. The Metadata is then fed to Oracle to work in conjunction with the web application server.

RTF to HTML Conversion

In current tests, the system generates around 9600 Rich Text Files. These need to be converted to HTML files, which are what the web server needs.

Other Processing

This step creates several specialized text files. Examples are the Program & Topic Pairs, and User Registration. These files are fed to Oracle databases for various purposes.

Actual Processing Log

This section provides information on how many documents are received, processed, and output from the Migration.

Processing Details

Preliminary Steps

1. Move final IFAP DBs from CD (provided by DOE), to a local drive.

2. Add All Docs view to all 39 IFAP DBs.

3. Add these special views to IFAP DBs:

   > In Regulations Compilations: copy view RegPartsAffected
   > In Federal Registers: copy views Reg Parts Affected Only

4. Remove these documents:

   > Regulations Compilations: Documents 21, 22, and 23. These 3 documents cause unpredictable errors, because of the size of the BodyOfText fields. They need to be processed manually.

5. Make sure all IsIncluded & IsMeta flags are set according to spreadsheet.

Main Processing

6. Now run this sequence of agents in all AFISA 1, AFISA 2, AFISA 3, AFISA 4, AFISA 5, AFISA 6, and AFISA 8 migration databases. For now, skip AFIS 7, which is actually IPOS 1, IPOS 2, and IPOS3.

   In the All Included Data In Sequence view

   > Delete All Docs in the view

   In the All Fields In Sequence view

   > Agent: Write Included Data In Sequence

   In the All Included Data In Sequence view:

   > Agent: 3a. Process RFT Part 1
   > Agent: 3b. Process RFT Part 3
   > Agent:3c. Combine Parts into NewRTF
   > Agent:3d. Remove Attachments
   > Agent:3e. Export RTF Files
   > Agent:4. Write Metadata text File

RTF to HTML Conversion

7. Use NT Explorer to display the /docs directory, where all RTF files are stored. Highlight 300 or 400 at a time, and feed to the DOX program box. The files will be converted to HTML, and written to the directory from where they came.

8. Break the IPOS Guidance Documents into the IPOSOnly1, IPOSOnly2, and IPOSOnly3 databases. IPOSOnly1 gets the 'IPO Docs' documents. IPOSOnly2 gets the 'Procedure Memos' documents. IPOSOnly3 gets the 'Program Review Guide' documents.

    Now run the above Agent sequence in each of the IPOS migration databases.

Other Processing

9. In the Program & Topics Pairs DB, run the Write All Program & Topics Pairs agent. This creates the AllProgramTopicsPairs.txt file in the /docs directory.

10. In the User Registration DB, run these 3 agents: 2. Write Data Files 1, Write Data Files 2, Write Data Files 3. This creates 6 files in /docs: UserReg1-3.dat, and UserReg1-3.schema.dat.

11. In the URL disposition database, run the Write Midas Hyperlinks agent. This creates the 4 MidasHyp_* files.

12. In the AFIS 6 database, run this agent:

    In the FedRegComp view, Agent: BuildFedREGCOmp

    In the AFIS 4 database, run these agents:

    In the RegPartsAffected #2 view, Agent: RPA#1
    In the RegPartsAffected #2 view, Agent: Write RegPartsAffected File

    Together, these 3 agents create the RegPartsAffected.dat file in /docs.

Final Processing Log

These are the results of a full processing run on October 27th and 28th.

AFIS #1

IFAP Notes Databases	# Docs in IFAP DB	Docs Migrated	# RTFs Created	#Lines Metadata	# Attach-ments / No Extension
Accrediting Agencies	28	28	28	28	3 – 0
Audit Guide	18	18	18	18	7 - 0
Campus-Based Publications	22	22	22	22	16 – 1
Conference & Presentation	69 *	69	69	69	186 – 0
The Blue Book	61	61	61	61	14 – 0
Comments	* This is the original confpres.nsf. The current version has many invalid pdf files.				

AFIS #2

IFAP Notes Databases	# Docs in IFAP DBs	Docs Migrated	# RTFs Created	#Lines Metadata	# Attach-ments / No Extension
Counselor's Handbook	155	155	155	155	37 – 0
Official Cohort Default Rates	138	138	138	138	159 – 2
Direct Loan Fact Sheet	12	12	12	12	11
Direct Loan School Guide	50	50	50	50	39 – 0
Direct Loan Newsletters	17	17	17	17	4
Comments					

AFIS #3

IFAP Notes Databases	# Docs in IFAP DB	Docs Migrated	# RTFs Created	#Lines Metadata	# Attach-ments / No Extension
Direct Loan Pamphlets	13	13	13	13	44 – 0
Direct Loan Tech Reference	44	44	44	44	74 – 12
Draft SFA Publications	19	19	19	19	17 – 0
EDConnect For Windows	16	16	16	16	11
EDE Technical Reference	33	33	33	33	21 – 0
Comments					

AFIS #4

IFAP Notes Databases	# Docs in IFAP DB	Docs Migrated	# RTFs Created	#Lines Metadata	# Attach- ments / No Extension
EFC Formula Book	12	12	12	12	10 – 0
FAFSA Instructions	13	13	13	13	18 – 0
Federal Registers	301	301	301	301	321 – 5
FFEL Variable Interest Rates	4	4	4	4	3 – 0
FFEL Special Allowance Rates	14	14	14	14	14 – 0
SFA Handbook	407	407	407	407	177 – 1
Comments					

AFIS #5

IFAP Notes Databases	# Docs in IFAP DB	Docs Migrated	# RTFs Created	#Lines Metadata	# Attach- ments / No Extension
FWS Resource Guide	16	16	16	16	0
1998 HEA Reauthorization	19	19	19	19	0
NSLDS Materials	84	84	84	84	68 – 0
SAR/ISIR Reference Materials	70	70	70	70	84 – 0
SSCR User's Guide	104	104	104	104	44 – 11
Training Guides	89	89	89	89	162 – 7
Verification Guide	115	115	115	115	7 – 0
Comments					

AFIS #6

IFAP Notes Databases	# Docs in IFAP DB	Docs Migrated	# RTFs Created	#Lines Metadata	# Attach- ments / No Extension
Regulations Compilations	4,569	4,569	4,569	4,569	47 – 3
Comments	This DB has been isolated to its own AFIS file because it contains many documents and takes a long time to process. Processing the RTF files is crashes the computer several times. This means we have to be very careful that the correct number of HTML files are created.				

AFIS #7

IFAP Notes Databases	# Docs in IFAP DB	Docs Migrated	# RTFs Created	#Lines Metadata	# Attach-ments / No Extension
IPOS 1 – IPOS Docs	4	4	4	4	36
IPOS 2 – Procedure Memos	8	8	8	8	
IPOS 3 – Program Review Guides	46	44	44	44	
Comments					

AFIS #8

IFAP Notes Databases	# Docs in IFAP DB	Docs Migrated	# RTFs Created	#Lines Metadata	# Attach-ments / No Extension
Announcements	1,030	1,030	1,099	1,030	98 – 6
Direct Load Bulletins	277	277	339	277	75 – 1
Dear Colleage Letters	704	704	1,132	704	229 – 4
Pell / RFMS Materials	92	92	92	92	32 – 1
Policy Bulletins	8	8	8	8	0
Comments	These are the databases containing Enclosures, which have to be processed separately. Essentially, the Enclosure RTFs must be processed before and in a different way than normal.				

Other Output Files

IFAP Notes Databases	Lines Output			
Program & Topic Pairs	7,732			
URL Disposition …				
… Notes Docs	794			
… Others	1,033			
User Registration	6,746			
Citations	4,490			
Comments				